Identification from data with periodically missing output samples

Ivan Markovsky [a], Mohammad Alsalti [b], Victor G. Lopez [b], and Matthias A. Müller [b]

[a] *I. Markovsky is with the Catalan Institution for Research and Advanced Studies (ICREA), Pg. Lluis Companys 23, Barcelona, Spain and the International Centre for Numerical Methods in Engineering (CIMNE), Gran Capitàn, Barcelona, Spain.*

[b] *Leibniz University Hannover, Institute of Automatic Control, 30167 Hannover, Germany*

Abstract

The identification problem in case of data with missing values is challenging and currently not fully understood. For example, there are no general nonconservative identifiability results, nor provably correct data efficient methods. In this paper, we consider a special case of periodically missing output samples, where all but one output sample per period may be missing. The novel idea is to use a lifting operation that converts the original problem with missing data into an equivalent standard identification problem. The key step is the inverse transformation from the lifted to the original system, which requires computation of a matrix root. The well-posedness of the inverse transformation depends on the eigenvalues of the system. Under an assumption on the eigenvalues, which is not verifiable from the data, and a persistency of excitation-type assumption on the data, the method based on lifting recovers the data-generating system.

Key words: system identification, missing data, behavioral approach, lifting.

1 Introduction

The standard setup in system identification is that the data consists of equidistantly sampled input/output variables of the to-be-identified system. If even a single variable is missing at one moment of time, the whole dataset can not be used by standard identification methods. Due to the time-series nature of the data, unless the missing variable appears in the first or the last sample, it can not be skipped. It should be estimated together with the system. This led to the development of specialized methods for identification of linear time-invariant systems from data with missing values, see, *e.g.*, [9,3,22,8,11,21,18,14]. Alternatively, the data can be split into two datasets—one before the missing sample and one after it. The two data sets should be fitted then by one model. For multiple missing variables, scattered over different moments of time, the data becomes fragmented into multiple short data sets. The worst distribution of the missing samples is periodic and the best is lumped missing data in the beginning or at the end. In the former case, the data is maximally fragmented into short sub-trajectories, while in the latter case the missing samples can be discarded, leaving one maximally long complete trajectory. Efficient methods that fit simultaneously multiple data sets are proposed in [17]. Applied to the problem of identification from data with missing values, however, they are not data efficient because of unexploited relations among the data sets.

The approach based on simultaneous estimation of the missing data and identification of the underlying data-generating system is data efficient, however, it leads to a nonconvex optimization problem. There are different heuristics in the literature that can be used for solving it. We mention the local optimization method based on the variables projections [18,13], methods using convex relaxations based on the nuclear norm [5,11], and the subspace method of [14]. Particularly interesting is the subspace method of [14] as it is noniterative and has theoretical guarantees to recover the data-generating system in case of exact data satisfying suitable assumptions. These assumptions, however, can not be checked a priori as they are not explicitly given as in the classical identification problem [23]. A modification of the method of [14] and a priori verifiable sufficient conditions on the input for identifiability are derived in [1]. The method of [14] and its modification in [1], however, are not data efficient, *i.e.*, there are cases when the system is identifiable but the methods can not identify it.

In this paper we consider the special case of linear time-

⋆ Corresponding author I. Markovsky.

Email addresses: imarkovsky@cimne.upc.edu (Ivan Markovsky), alsalti@irt.uni-hannover.de (Mohammad Alsalti), lopez@irt.uni-hannover.de (Victor G. Lopez), mueller@irt.uni-hannover.de (Matthias A. Müller).

invariant system identification from exact data with periodically missing output samples and fully specified inputs. As explained above, the periodic distribution of the missing values is the worst case in the sense that it leads to a maximal fragmentation of the data. Indeed, if P is the period of the missing values, the longest complete sub-trajectory of the data trajectory has length $P-1$. Note that classical system identification methods, such as the N4SID and MOESP methods [19,20], require trajectories with length at least equal to the lag ℓ of the system plus one in order to form a data matrix of sufficient size. Therefore, for $P \leq \ell+1$ classical identification methods can not recover the system using the splitting method. In this setup, the methods of [14] and [1] can recover the data-generating system when $P = \ell+1$ and the order of the system is larger than the number of outputs, but not when $P < \ell+1$. This made us hypothesize that the system is not identifiable when $P < \ell+1$. The present paper rejects the hypothesis, providing a method that recovers the system for any periodicity P of the missing output samples as well as any number of outputs. Moreover, as few as one output sample per period may be given. Thus, the new method is more data efficient than the one of [14].

The novel idea at the core of the proposed method is to use a *lifting operation* which subsamples the data at a rate P, then concatenates the P input/output samples over one period into a single sample. Provided that the data-generating system is linear time-invariant and the data is exact, the lifted data is also a trajectory of a linear time-invariant system, called the *lifted system*. Selecting the given elements from the outputs of the lifted system results in a complete trajectory from which the lifted system's dynamics can be identified by standard system identification methods. The inverse transformation from the lifted system to the original system is well-posed provided that the eigenvalues of the original system (*i.e.*, the eigenvalues of the A matrix in a minimal state-space representation) are in the sector

$$\mathscr{C}_P := \{\, \lambda \in \mathbb{C} \mid -\pi/P < \angle\lambda < \pi/P \}$$

where $\angle\lambda$ is the angle of λ and the lifted system has no real negative or zero eigenvalues [2]. The condition on the eigenvalues of the original system is not verifiable from the data and restrict the applicability of the method.

Section 2 introduces the lifting operation by its action on a trajectory. Applied on a system, it results in a system that preserves linearity and time-invariance. Theorem 1 shows how the lifting operation changes the complexity of the system, which is defined as the triple: number of inputs, lag, and order. Theorem 2 shows the effect of the lifting operation on an input/state/output representation of the system. Based on the results in Section 2, Section 3 presents the proposed identification method for identification with periodically missing output samples. The key step of the method is the inverse transformation of the identified lifted system to the original system. The operation requires computing the P-th root of a matrix. The main result is stated in Theorem 4, which gives conditions under which the method recovers the true data-generating system, *i.e.*, an identifiability result for identification with periodically missing output samples. Section 4 empirically validates and illustrates the method.

2 Methodology: the lifting operation

We use the behavioral approach, where a dynamical system $\mathscr{B}$ is defined as a set of trajectories. Thus, $w \in \mathscr{B}$ means that w is a trajectory of the system $\mathscr{B}$. The restriction of the trajectory w and the system $\mathscr{B}$ to the finite interval $[1,T]$ is denoted by $w|_T$ and $\mathscr{B}|_T$, respectively. The class of linear time-invariant systems with q variables is denoted by $\mathscr{L}^q$ and the class of linear time-invariant systems with bounded complexity is denoted by $\mathscr{L}^q_{m,\ell,n}$, where m, ℓ, n are upper bounds of the number of inputs, lag, and order, respectively. A $\mathscr{B} \in \mathscr{L}^q_{m,\ell,n}$ admits an *input/state/output representation*

$$\mathscr{B} = \mathscr{B}_{\mathrm{ss}}(A,B,C,D,\Pi) := \{\, \Pi \left[\begin{smallmatrix} u \\ y \end{smallmatrix}\right] \mid \text{there is } x \in (\mathbb{R}^n)^{\mathbb{N}}, \text{ such that } \sigma x = Ax + Bu,\ y = Cx + Du \,\}, \quad \text{(I/S/O)}$$

where σ is the shift operator, defined by $(\sigma w)(t) := w(t+1)$, and $\Pi \in \mathbb{R}^{q\times q}$ is a permutation matrix. The number of inputs, lag, and order of a system $\mathscr{B}$ are denoted by $\mathbf{m}(\mathscr{B})$, $\mathbf{l}(\mathscr{B})$, and $\mathbf{n}(\mathscr{B})$, respectively.

The *Hankel matrix* with L block rows, constructed from the trajectory $w = \big(w(1),\dots,w(T)\big) \in (\mathbb{R}^q)^T$, is

$$\mathscr{H}_L(w) := \begin{bmatrix} w(1) & w(2) & \cdots & w(T-L+1) \\ w(2) & w(3) & \cdots & w(T-L+2) \\ \vdots & \vdots & & \vdots \\ w(L) & w(L+1) & \cdots & w(T) \end{bmatrix}.$$

For a trajectory $w \in \mathscr{B}|_T$ and a natural number $P \leq T$, we define the *lifted trajectory* $w' \in (\mathbb{R}^{qP})^{T'}$ as

$$\begin{aligned} w' &= \big(w'(1), w'(2), \dots, w'(T')\big) \\ &= \underbrace{\left(\begin{bmatrix} w(1) \\ \vdots \\ w(P) \end{bmatrix}, \begin{bmatrix} w(P+1) \\ \vdots \\ w(2P) \end{bmatrix}, \dots, \begin{bmatrix} w\big((T'-1)P+1\big) \\ \vdots \\ w(T'P) \end{bmatrix} \right)}_{\mathrm{lift}_P(w)}, \end{aligned}$$

where $T' := \lfloor T/P \rfloor$ is the largest integer smaller than or equal to T/P. The lifting operator lift_P is used for identification of periodically time-varying systems [16]. In this paper, we use it for identification from data with periodically missing output samples. The matrix $\mathrm{lift}_P(w)$ coincides with the Page matrix of w with depth P [4].

Acting on a system $\mathscr{B}$, the lifting operator creates a system

$$\mathscr{B}' = \text{lift}_P(\mathscr{B}) := \{\, \text{lift}_P(w) \mid w \in \mathscr{B} \,\}.$$

The following theorem asserts that if $\mathscr{B}$ is linear time-invariant, $\text{lift}_P(\mathscr{B})$ is also linear time-invariant. The theorem also characterizes the complexity of $\text{lift}_P(\mathscr{B})$ in terms of the complexity of $\mathscr{B}$.

Theorem 1 (Complexity of lift$_P(\mathscr{B})$) *Let $\mathscr{B} \in \mathscr{L}^q_{m,\ell,n}$ and $P \in \mathbb{N}$. Then, lift$_P(\mathscr{B}) \in \mathscr{L}^{Pq}_{m',\ell',n'}$, where*

$$m' = mP, \quad \ell' = \lceil \ell/P \rceil, \quad \textit{and} \quad n' = n$$

($\lceil \ell/P \rceil$ is the smallest integer larger than or equal to ℓ/P).

PROOF. By construction $\mathscr{B}|_{LP} = \mathscr{B}'|_L$ for any $L, P \in \mathbb{N}$. Using this fact and the formula (see [15, Corollary 5])

$$\dim \mathscr{B}|_t = \mathbf{m}(\mathscr{B})t + \mathbf{n}(\mathscr{B}), \text{ for all } t \geq \mathbf{l}(\mathscr{B})$$

applied to $\mathscr{B} \in \mathscr{L}^q_{m,\ell,n}$, we have

$$\dim \mathscr{B}|_{LP} = \dim \mathscr{B}'|_L = mPL + n, \quad \text{for } PL \geq \ell. \qquad (*)$$

Using [15, Corollary 5] applied to $\mathscr{B}' \in \mathscr{L}^{qP}_{m',\ell',n'}$, we have

$$\dim \mathscr{B}'|_L = m'L + n', \quad \text{for } L \geq \ell'. \qquad (**)$$

The complexity of $\mathscr{B}'$ follows by comparison of $(*)$ and $(**)$.

In what follows we assume that the system $\mathscr{B}$ admits an input/output partitioning

$$w = \begin{bmatrix} u \\ y \end{bmatrix}, \quad \text{where } u(t) \in \mathbb{R}^m \text{ and } y(t) \in \mathbb{R}^p,$$

with $p := q - m$. Equivalently, in the input/state/output representation (I/S/O) we assume $\Pi = I$ (and skip it from the notation). The following result shows a particular input/state/output representation of the lifted system $\text{lift}_P(\mathscr{B})$ that manifest a link to the parameters of an input/state/output representation of $\mathscr{B}$.

Theorem 2 (State-space representation of lift$_P(\mathscr{B})$) *Let $\mathscr{B} = \mathscr{B}_{\text{ss}}(A,B,C,D)$ and $P \in \mathbb{N}$. Then, the lifted system has a state-space representation*

$$\textit{lift}_P(\mathscr{B}) = \mathscr{B}_{\text{ss}}(A',B',C',D'), \quad \textit{with}$$
$$A' = A^P, \quad B' = \begin{bmatrix} A^{P-1}B & \cdots & AB & B \end{bmatrix},$$
$$C' = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{P-1} \end{bmatrix}, \quad D' = \begin{bmatrix} D & & & \\ CB & D & & \\ \vdots & \ddots & \ddots & \\ CA^{P-1}B & \cdots & CB & D \end{bmatrix}. \qquad \text{(SS)}$$

Moreover, if $\mathscr{B}_{\text{ss}}(A,B,C,D)$ is minimal, then $\mathscr{B}_{\text{ss}}(A',B',C',D')$ is also minimal.

PROOF. For any trajectory $w' = \begin{bmatrix} u' \\ y' \end{bmatrix} \in \mathscr{B}'$ of the lifted system and for any $t \in \mathbb{N}$, we have that $w'(t) \in \mathscr{B}_{\text{ss}}(A,B,C,D)|_P$. Therefore,

$$\begin{aligned} x\big(tP+1)\big) = A^P x\big((t-1)P+1\big) \\ + \begin{bmatrix} A^{P-1}B & \cdots & AB & B \end{bmatrix} u'(t) \end{aligned}$$

and

$$y'(t) = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{P-1} \end{bmatrix} x\big((t-1)P+1\big) +$$
$$\begin{bmatrix} D & & \\ CB & D & \\ \vdots & \ddots & \ddots \\ CA^{P-2}B & \cdots & CB \ D \end{bmatrix} u'(t).$$

The result follows by comparing the above equations and

$$\begin{aligned} \sigma^P x &= A'x + B'u' \\ y' &= C'x + D'u'. \end{aligned}$$

If $\mathscr{B}_{\text{ss}}(A,B,C,D)$ is minimal, $\dim(A) = \mathbf{n}(\mathscr{B})$ and by Theorem 1, $\mathbf{n}(\mathscr{B}') = \mathbf{n}(\mathscr{B})$. The minimality of $\mathscr{B}_{\text{ss}}(A',B',C',D')$ follows from $\dim(A') = \dim(A) = \mathbf{n}(\mathscr{B}')$.

3 Identification with periodically missing outputs

In this section, we consider the problem of identification from data with periodically missing output samples. Let $\mathscr{B} \in \mathscr{L}^q_{m,\ell,n}$ be the data-generating system and $w_{\text{d}} \in \mathscr{B}|_{T_{\text{d}}}$ be the data trajectory with missing output samples with period P. As explained in the introduction, the problem is challenging when $P \leq \mathbf{l}(\mathscr{B}) + 1$. Moreover, for the case $P < \mathbf{l}(\mathscr{B}) + 1$ there are no alternative methods that solve this problem. Next, we show how the problem can be solved using the lifting operation and a back transformation derived from (SS). Since the method converts the problem of identification with missing data to an identification with complete data, any method can be used for the latter. The method proposed for the missing data estimation problem is provably correct under given conditions and can deal with the situation of $\mathbf{l}(\mathscr{B}) = 1$ as well as up to $P - 1$ missing output samples per period.

For simplicity assume that only one sample is given per period and it is the first one. Let $w'_{\rm d} := \mathrm{lift}_P(w_{\rm d})$ and let $w''_{\rm d}$ be the trajectory obtained from $w'_{\rm d}$ by removing the missing elements. It follows from Theorems 1 and 2 that $w''_{\rm d}$ is a trajectory of a system $\mathscr{B}'' \in \mathscr{L}^{mP+p}_{mP,\ell'',n}$ for some ℓ'' with a state-space representation

$$\begin{aligned}\mathscr{B}'' &= \mathscr{B}_{\rm ss}(A'',B'',C'',D'') \\ &:= \mathscr{B}_{\rm ss}\Big(A^P,\ \begin{bmatrix}A^{P-1}B & \cdots & AB & B\end{bmatrix},\ C,\ \begin{bmatrix}D & 0 & \cdots & 0\end{bmatrix}\Big).\end{aligned}$$

Under the generalized persistency of excitation [15]

$$\operatorname{rank} \mathscr{H}_{\ell''+1}(w''_{\rm d}) = mP(\ell''+1) + \mathbf{n}(\mathscr{B}), \qquad \text{(GPE)}$$

the system $\mathscr{B}''$ can be recovered back from $w''_{\rm d}$ by computing the *most powerful unfalsified model* $\mathscr{B}_{\rm mpum}(w''_{\rm d})$.

Remark 3 (Upper bound on the lag of $\mathscr{B}''$) *An upper bound $\ell'' \leq P\ell$ can be used in (GPE) as well as in the following results: Theorems 4 and 6 and ($T_{\min}$). Moreover, under certain assumptions [10], it can be guaranteed that $\ell'' = \ell$.*

A representation of $\mathscr{B}_{\rm mpum}(w''_{\rm d})$ can be computed in practice using standard identification methods. This leads to the following procedure for system identification from data with periodically missing output samples:

(1) find $\mathscr{B}_{\rm mpum}(w''_{\rm d}) := \mathscr{B}_{\rm ss}(A'',B'',C'',D'')$,
(2) compute the P-th matrix root $\widehat{A}$ of A'',
(3) let $\widehat{B}$ be the last block-element of B'', $\widehat{C} = C''$, and $\widehat{D}$ the first block-element of D''.

The identified model $\widehat{\mathscr{B}}$ is then given by $\widehat{\mathscr{B}} := \mathscr{B}_{\rm ss}(\widehat{A},\widehat{B},\widehat{C},\widehat{D})$.

Our main result is the following theorem that gives conditions under which the model identified by the method above coincides with the true data-generating system.

Theorem 4 (Identifiability conditions) *Assuming that*

(1) $\mathscr{B}''$ is identifiable from $w''_{\rm d}$, i.e., *(GPE) holds,*
(2) $\mathscr{B}''$ has no real negative or zero eigenvalues, and
(3) the eigenvalues of $\mathscr{B}$ are in the sector $\mathscr{C}_P$,

the identified model is exact, i.e., $\mathscr{B}_{\rm ss}(\widehat{A},\widehat{B},\widehat{C},\widehat{D}) = \mathscr{B}$.

PROOF. By Assumption 1 and [15, Theorem 17], $\mathscr{B}''$ is identifiable, *i.e.*, $\mathscr{B}_{\rm mpum}(w''_{\rm d}) = \mathscr{B}''$. By Assumption 2, A'' has a unique P-th root $\widehat{A}$ with eigenvalues in $\mathscr{C}_P$, see [6, Theorem 4.8] and [2]. By Assumption 3, $\widehat{A} = A$ up to a similarity transformation.

The conclusion $\mathscr{B}_{\rm ss}(\widehat{A},\widehat{B},\widehat{C},\widehat{D}) = \mathscr{B}$ follows from Theorem 2, *i.e.*, there is a representation $\mathscr{B}_{\rm ss}(\widehat{A},\widehat{B},\widehat{C},\widehat{D})$ of $\mathscr{B}$, such that the corresponding lifted state-space representation matches the computed model parameters A'',B'',C'',D'' from the data $w''_{\rm d}$. If $\mathscr{B}_{\rm ss}(A'',B'',C'',D'')$ is minimal, then $\mathscr{B}_{\rm ss}(\widehat{A},\widehat{B},\widehat{C},\widehat{D})$ is also minimal.

Remark 5 (Minimal number of samples for identifiability) *For Assumption 1 to be satisfied, the number of samples T should be at least*

$$T_{\min} := P\big((mP+1)(\ell''+1) + \mathbf{n}(\mathscr{B}) - 1\big). \qquad (T_{\min})$$

The alternative identifiability result stated in the following theorem is equivalent to the fundamental lemma [23, Lemma 1] for the problem of system identification with periodically missing output samples. The order of persistency of excitation $\mathrm{PE}(u)$ of the signal $u \in (\mathbb{R}^m)^T$ is the maximal $L \in \mathbb{N}$, for which the Hankel matrix $\mathscr{H}_L(u)$ is full row-rank.

Theorem 6 *Assuming that*

(1) the data-generating system $\mathscr{B}$ is controllable,
(2) $\mathrm{PE}(u') = \ell'' + \mathbf{n}(\mathscr{B}) + 1$, and
(3) Assumptions 2 and 3 of Theorem 4 hold,

the identified model is exact, i.e., $\mathscr{B}_{\rm ss}(\widehat{A},\widehat{B},\widehat{C},\widehat{D}) = \mathscr{B}$.

PROOF. The result follows by replacing the identifiability assumption (Assumption 1) in Theorem 4 by the sufficient identifiability conditions of [23].

4 Numerical examples

This section illustrates the effectiveness of the proposed method on a numerical example. The results are reproducible and the code is available as a supplementary material from

```
https://imarkovs.github.io/mpum-md.tgz
```

The simulation parameters are: the number of inputs m, number of outputs p, the order n of the data generating system $\mathscr{B}$, the period P of the missing values, and the number of samples $T_{\rm d}$:

```
%% simulation setup
m = 1; p = 1; q = m + p; n = 5; P = 2; Td = P * 100;
```

The data generating system $\mathscr{B}$ is a randomly selected.

```
%% random stable system
B = drss(n, p, m);
```

The data $w_{\rm d}$ is generated as a random trajectory of $\mathscr{B}$ with periodically missing outputs, where only one output sample per period is given:

```
%% generate data
ud = rand(Td, m); yd0 = lsim(B, ud);
yd = NaN(size(yd0));
yd(1:P:end, :) = yd0(1:P:end, :);
wd = [ud yd];
```

The method described in Section 3 and implemented in the function sysid_pmo is applied on the data w_d. The result $\widehat{\mathscr{B}}$ is validated by computing the relative H_2-error

$$e := \|H - \widehat{H}\| / \|H\|,$$

where H is the transfer function of $\mathscr{B}$ and $\widehat{H}$ is the transfer function of $\widehat{\mathscr{B}}$.

```
%% apply the method
Bh = sysid_pmo(wd, m, n, P);
error = norm(B - Bh) / norm(B)
```

Provided that Assumptions 2 and 3 of Theorem 4 are satisfied (which depends on the random system generated by drss), a typical result for the relative H_2-error e is of the order of 10^{-10}. This empirical fact confirms Theorem 4, *i.e.*, that the method recovers the data-generating system. Alternative methods, *e.g.*,

- the method of [12] (implemented in the misdata function of Matlab's System Identification Toolbox),
- the structured low-rank approximation method of [18] (implemented in the ident function of [13]),
- the nuclear norm method [5,11] (implemented using the CVX package [7]), and
- the subspace method of [14],

fail irrespective of the choice of the simulation parameters and the system generated by drss.

5 Conclusions

We conjecture that the method proposed in the paper is data efficient for the problem of periodically missing output samples. Relaxing the assumptions and investigating the data efficiency of the method is left for future work.

Acknowledgments

The research leading to these results has received funding from the Catalan Institution for Research and Advanced Studies (ICREA), the Fond for Scientific Research Vlaanderen (FWO) project G033822N, and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant No 948679).

References

[1] M. Alsalti, I. Markovsky, V. G. Lopez, and M. A. Müller. Data-based system representations from irregularly measured data, 2023.

[2] D.A. Bini, N.J. Higham, and B. Meini. Algorithms for the matrix pth root. *Numerical Algorithms*, pages 349–378, 2005.

[3] P. Broersen, S. De Waele, and R. Bos. Autoregressive spectral analysis when observations are missing. *Automatica*, 40(9):1495–1504, 2004.

[4] A. Damen, P. Van den Hof, and A. Hajdasinski. Approximate realization based upon an alternative to the Hankel matrix: the Page matrix. *Systems & Control Lett.*, 2:202–208, 1982.

[5] M. Fazel. *Matrix rank minimization with applications*. PhD thesis, Elec. Eng. Dept., Stanford University, 2002.

[6] J. Gallier. Logarithms and square roots of real matrices. arXiv: 0805.0245, 2008.

[7] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming. stanford.edu/~boyd/cvx, 2017.

[8] A. Hansson and R. Wallin. Maximum likelihood estimation of gaussian models with missing data—Eight equivalent formulations. *Automatica*, 48(9):1955–1962, September 2012.

[9] A. Isaksson. Identification of ARX-models subject to missing data. *IEEE Trans. Automat. Contr.*, 38(5):813–819, May 1993.

[10] Isabelle Krauss, Victor G. Lopez, and Matthias A. Müller. Sample-based observability of linear discrete-time systems. In *IEEE 61st Conf. on Decision and Control*, pages 4199–4205, 2022.

[11] Z. Liu, A. Hansson, and L. Vandenberghe. Nuclear norm system identification with missing inputs and outputs. *Systems & Control Lett.*, 62:605–612, 2013.

[12] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, 1999.

[13] I. Markovsky. A software package for system identification in the behavioral setting. *Control Eng. Practice*, 21:1422–1436, 2013.

[14] I. Markovsky. On the most powerful unfalsified model for data with missing values. *Systems & Control Lett.*, 95:53–61, 2016.

[15] I. Markovsky and F. Dörfler. Identifiability in the behavioral setting. *IEEE Trans. Automat. Contr.*, 68:1667–1677, 2023.

[16] I. Markovsky, J. Goos, K. Usevich, and R. Pintelon. Realization and identification of autonomous linear periodically time-varying systems. *Automatica*, 50:1632–1640, 2014.

[17] I. Markovsky and R. Pintelon. Identification of linear time-invariant systems from multiple experiments. *IEEE Trans. Signal Process.*, 63(13):3549–3554, 2015.

[18] I. Markovsky and K. Usevich. Structured low-rank approximation with missing data. *SIAM J. Matrix Anal. Appl.*, 34(2):814–830, 2013.

[19] P. Van Overschee and B. De Moor. *Subspace identification for linear systems: Theory, implementation, applications*. Kluwer, Boston, 1996.

[20] M. Verhaegen and P. Dewilde. Subspace model identification, Part 1: The output-error state-space model identification class of algorithms. *Int. J. Contr.*, 56:1187–1210, 1992.

[21] R. Wallin and A. Hansson. Maximum likelihood estimation of linear SISO models subject to missing output data and missing input data. *Int. J. Control*, 87(11):2354–2364, 2014.

[22] Y. Wang, J. Li, and P. Stoica. *Spectral Analysis of Signals: The Missing Data Case*. Morgan and Claypool, 2006.

[23] J. C. Willems, P. Rapisarda, I. Markovsky, and B. De Moor. A note on persistency of excitation. *Systems & Control Lett.*, 54(4):325–329, 2005.